Effect of Task on Acoustic Correlates of Stress

Implications for Research Methodology

Kristin A. King*[1], Michael Rastatter[2]

[1]Department of Audiology and Speech Pathology, University of Tennessee – Health Science Center, Knoxville, Tennessee 37996, USA

[2]Department of Communication Sciences and Disorders, East Carolina University, Greenville, NC 27858, USA

*[1]kking29@uthsc.edu; [2]mrastatter@ecu.edu

Abstract

When addressing acoustic parameters of speech, much of the previous research has focused on fundamental frequency (F0), duration, and intensity as the primary acoustic characteristics within a prosodic structure. The findings from these studies are limited in relation to natural speech as they tend to focus on structured, controlled tasks with best productions. This study addresses questions related to the effect of task on the acoustic correlates of stress in sentences, in addition differences in productions were investigated when following an instruction versus following a question. Fifteen speakers of general American English were recorded in two speaking tasks. Tasks used were: (1) a detailed instruction as to what should be produced and (2) a question, without any instruction, to elicit responses. Measurements of fundamental frequency (F0), amplitude, duration, and total duration were taken in all contexts. Findings revealed significant differences between tasks for all acoustic parameters measured: F0, amplitude, and duration. The pattern for the magnitude of the change also varied between the tasks. These findings suggested that the natural variation observed in spoken language is reduced with directed tasks and that an individual's production of stress changes dependent on the task. Implications for past and future research have been discussed.

Keywords

Acoustic; Prosody; Intonation; Methodology; Speech; Fundamental Frequency; Amplitude; Duration

Introduction

Prosody plays a central role in human communication. During speech acts, the intended message is not conveyed just by words but by varying the intonational or suprasegmental attributes of the message, such as prosody. Much research has been conducted to identify and quantify the contributing acoustic parameters to intonation. Prosody and intonational patterns effect how speech is perceived by others. Defining prosody involves identifying the suprasegmental features inherent in a message, which generally includes pitch, rate, intensity, and duration (Cutler, Dahan, and van Donselaar, 1997). Much research has focused on the role of fundamental frequency (F0), duration, and intensity in conveying linguistic and nonlinguistic information. However, most of the research to date has studied these acoustic parameters in the context of a structured task (one that uses repetition of a speech production sample) rather than being applied to "natural" speech productions. To date, however, a paucity of data exists in these areas based on natural speaking environments.

Acoustic cues used for the production of intonation differentiate stress within words and phrases and may vary the meaning of the message. Lieberman (1960) indicated that typically two or three acoustic parameters (frequency, amplitude, or duration) show an increase in magnitude during production of sentence stress. Based on these three parameters, it has been postulated that fundamental frequency (F0) and duration are the most salient parameters for marking stress (Cooper, Eady, and Mueller, 1985; Fry, 1958; Howell, 1993). Additional research has indicated that stress conveys meaning for both emotional and linguistic structures (Monnot, Orbello, Riccardo, Sikka, and Ross, 2003; Pell, 2007; Shtyrov, Pihko, and Pulvermuller, 2005; Snow, 2000).

Researchers have addressed the question of which acoustic parameters convey stress by analyzing various acoustic attributes within word and sentence productions. Previous research has been conducted to study prosody in various contexts with both normal and disordered populations, such as dysarthria (Patel, 2003), children with developmental apraxia of speech and phonological disorders (Munson, Bjorum, and Windsor, 2003), the effects of age and hormones on various suprasegmental features of speech (Muerer, Wender, Von Eye Corleta, and Capp, 2004), and finally Chinese Mandarin speakers with English as a second language (Chen, Robb, Gilbert and Lerman, 2001). While each of theses studies provided valuable information relative to suprasegmental productions, cross-comparisons among the results to develop a

model of production are difficult since each study uses a different research methodology, each study employing various modeling techniques.

As noted, much of the research has involved the use of a directed task to produce a set sentence or question. Research methods also have involved a selection bias in that only the best productions were utilized in the studies. An inherent difficulty with this method of research is the generalization of the findings to natural speech. A small body of past research had recognized these difficulties and attempted to measure the acoustic parameters with a consideration for natural speech productions. In these studies, the authors elicited sentence production by having the participants respond to a question (Cooper et al, 1985; Eady and Cooper, 1986). However, the task was repeated as needed with the experimenter judging the production and requiring repetition if it did not appear to exhibit the targeted stress; therefore, the selection bias eliminated the natural context of the productions analyzed.

Another consideration is the influence of the task instruction, and its effect on the productions. Ferguson (2004) addressed this question by asking whether or not the instruction given to the speaker would affect the perception of the task and therefore the production. All tokens were recorded following a given instruction to use either a clear or conversational tone. The participants also were given the opportunity to practice the sentences and provide feedback by the researcher as to the quality of the productions. Although the findings indicated a significantly higher rating for the “clear” speech productions, acoustical analysis was needed to address what specific parameters differentiated one task from the other. FLiu, Rio, Bradlow, and Zeng (2004) provided acoustical analysis of a similar task. In both studies, their analyses indicated that duration and amplitude increased with the instruction to use “clear” speech; and frequency was not analyzed. However, this research methodology of feedback and best production to improve quality, further removes the speech sample from a natural production.

While many of these research methodologies have been conducted to study the acoustic and perceptual correlates of stress in words, the tasks used have been structured and the responses selected for best productions only. In other words, the research conclusions may not be generalizable to natural speech. Some inherent questions arise with the past research methodology: are the acoustic correlates of stress the same for both structured tasks and natural speech productions? Is there a difference in the acoustical characteristics (peak frequency, peak amplitude, and duration) of the stressed word when produced following the instruction to stress the word versus the natural context of answering a question? What are the acoustic characteristics that change when stressing a target word? Do all three parameters change or some combination of them? This study attempts to answer these questions. Our hypotheses are: (1) the acoustic correlates of stress (peak frequency, amplitude, and duration) will be different between the tasks and (2) task via a given instruction will elicit less variation in peak frequency, amplitude and duration as compared to a natural production following a question.

Experimental Procedure

Participants

Participants included fifteen male speakers of general American English between the ages of 27 and 35 years (M = 30.94, SD = 3.77). Male only speakers were used to eliminate known gender differences in frequency, amplitude, and duration of speech patterns. No history of speech or language impairments was reported on a case history formor evident in responses during an initial interview, which was conducted by a licensed and certified speech-language pathologist. The participants also did not have any background or training in speech, theatre, or voice.

Measures

Individual utterances of the phrase “Bev loves Bob” were recorded in a quiet room with a microphone (unidirectional Logitech B130A USB Desktop microphone) placed approximately 30 cm from the individual’s lips with an orientation of 0° azimuth and -10° altitude. The microphone output was line-fed to a PowerMac G-4 computer (M5183) for digital input via the USB port. The speech samples were digitally recorded at a sampling rate of 44.1 kHz and quantized at 16 bit.

Procedure

Each participant produced a total of 30 recorded utterances of the phrase “Bev loves Bob,” under two conditions. In the first condition, the utterances were recorded following the instruction to “stress the word ______”(Bev, loves, or Bob). Five productions with each target stressed word were recorded, for a total of 15 recordings.

The second condition consisted of recordings produced following the presentation of a question to elicit the stressed word (ie: Who loves Bob?). Three different questions were asked so that all 3 words (Bev, Loves, Bob) were targeted individually. The 3 words were targeted five times each; again, for a total of 15 productions. No instructions or qualifications were given to the participants for their production following a question. Productions were targeted to be natural based on their perception of the question and the targeted appropriate response. The first five productions were saved for later analysis.

The two conditions were counter-balanced for each participant. Under each condition, the targeted responses were randomized for the target words.

Analysis

Analysis of Speech

The speech analysis program, Signalyze 3.1.2, was used to produce amplitude waveforms of the recorded samples and for a Fast Fourier Transformation for spectral acoustical analysis of the signals. Each recorded phrase was measured within Signalyze 3.1.2 to obtain peak frequency, peak amplitude, and the duration of each word both when stressed and unstressed. Peak frequency was obtained following a pitch extraction. The peak amplitude was obtained following an RMS envelope extraction to produce an amplitude waveform. Duration was measured in milliseconds from the initial point to the end point of voicing, as observed within a spectrogram. An overall total duration was measured from the initial voicing point of "Bev" to the end point in "Bob."

Intra-judge Reliability

Twenty percent of the data were reanalyzed in an effort to provide an index of intrajudge reliability. Intrajudge reliability as measured by Pearson Product correlation, was 1.00 ($p<.001$) for peak frequency, peak amplitude, duration of word length, and for total duration. Intrajudge reliability, as measured by percent of agreement, revealed peak frequency was 99.3%, peak amplitude 98.6%, duration 97.7%, and total duration 98.35%.

Statistical Analysis

To address if the acoustic parameters would be different between the tasks, a linear mixed model analysis was used to analyze each of the dependent variables: peak frequency, peak amplitude, and duration. This model included random effects for each participant with a variance components approach for modeling the error (Singer, 1998, unpublished data). This approach produced an analysis similar to a repeated measures analysis of variance. The within subject factors of task (2 levels) and word (3 levels) were included in a factorial design. This design allowed for the differences between the task and word levels to be addressed. Contingency tables were used to consider the variations present for tasks and words.

Results

Mean values for peak frequency, peak amplitude, duration, and total duration were obtained from each participants' 30 productions of the phrase "Bev loves Bob." Means, standard deviations, and coefficient of variations are presented in Table 1.

Perceptual Judgment Task

After acoustical measurements were obtained, all recordings were submitted to a perceptual judgment task. The recordings were presented to a group of ten naïve listeners. The listeners judged which word in the recorded phrase was stressed. The group correctly identified Bev as the stressed word in 92.2% of the trials, loves in 93.5% of the trials, and Bob in 95.5% of the trials.

Peak Frequency

Analysis indicated a statistically significant main effect for word (F 2, 111= 36.75, $p <.001$) when "Bev" was the target stressed word, meaning the peak frequency was significantly higher as compared to "loves" or "Bob," but no effect for the task (target word elicited by instruction or natural context) was observed and no interaction of task by word.

Statistically significant main effects for word (F2, 111= 61.24, $p <.001$) and task (F1,111= 6.501, $p = .012$) and an interaction for word by task (F2, 111= 3.184, $p = .045$) were indicated, when "loves" was the stressed word. These results indicated that peak frequency for "loves" was significantly higher as compared to the other words and was dependent on task (target word elicited by instruction or natural context question). Peak frequency of "loves" was significantly higher when a instruction to stress the word was given versus when occurring in a natural context. An interaction of word by task also occurred.

Analysis of "Bob," when targeted, indicated a statistically significant main effect for word (F 2, 111= 11.46, $p <.001$) but no effect for task and no interaction of task by word.

TABLE 1 MEAN, STANDARD DEVIATION, AND COEFFICIENT OF VARIATION FOR PEAK FREQUENCY, PEAK AMPLITUDE, DURATION FOR EACH STRESSED WORD AND TASK

	BEV			LOVES			BOB		
	Mean	SD	Coefficient of Variation	Mean	SD	Coefficient of Variation	Mean	SD	**Coefficient of Variation**
INSTRUCTION									
Peak Frequency	163.55	33.28	20.3	170.2	31.74	18.6	132.65	32.47	**24.5**
Peak Amplitude	71.19	3.66	5.1	71.09	3.12	4.4	66.12	3.23	**4.9**
Duration	369.96	89.24	24.1	381.46	82.92	21.7	370.99	31.37	**8.5**
QUESTION									
Peak Frequency	158.90	37.17	23.4	148.35	38.52	26.0	135.55	32.691	**24.1**
Peak Amplitude	68.01	2.26	3.3	66.10	5.57	8.4	63.29	5.10	**8.1**
Duration	256.41	37.28	14.5	359.64	91.56	27.1	364.04	34.4	**9.4**

Peak Amplitude

The analysis indicated statistically significant main effects for word ($F2, 111= 5.027$, $p = .008$) and task ($F1,111=19.34$, $p < .001$) but no interaction, indicating that peak amplitude was significantly higher for "Bev" when stressed as compared to the other words. Peak amplitude also was significantly higher when the manner of elicitation was a given instruction versus the natural context of question.

Statistically significant main effects for word ($F2, 111= 28.25$, $p <.001$) and task ($F1,111=37.96$, $p < .001$) and a marginally significant interaction ($F2, 111= 2.764$, $p =.06$) were indicated when measuring peak amplitude and "loves" was stressed. These results revealed that peak amplitude was significantly higher for "loves" when stressed as compared to the other words. Peak amplitude also was significantly higher when the manner of elicitation was a given instruction versus the natural context of a question. An interaction of word by task also was marginally significant.

Analysis of "Bob," when targeted, indicating a statistically significant main effect for word ($F 2, 111= 14.959$, $p <.001$) and for task ($F 2, 111= 19.25$, $p <.001$) but no interaction. The peak amplitude for "Bob" was significantly higher as compared to other words and was higher when produced following an instruction versus the natural context of answering a question.

Duration

The analysis indicated statistically significant main effects for word ($F2, 111= 20.19$, $p <.001$) and task ($F1,111=23.84$, $p < .001$) and an interaction for word by task ($F2, 111= 14.15$, $p <.001$), indicating that duration was significantly longer for "Bev" when stressed as compared to the other words and was dependent on task (target word elicited by instruction or question). Duration of "Bev" was significantly longer following an instruction to stress the word than when occurring in a natural context.

Statistically significant main effects for word ($F2, 111= 21.837$, $p <.001$) and task ($F1,111=18.571$, $p < .001$), but no interaction, were indicated when analyzing the duration of "loves." In other words, duration was significantly longer for "loves" when stressed, as compared to the unstressed word. The duration of "loves" also was longer dependent on the manner of elicitation. Duration of "loves" was significantly longer following the given instruction to stress the word rather than when occurring in a natural context (following a question).

When analyzing duration of "Bob," statistically significant main effects for word ($F2, 111= 40.027$, $p <.001$) were observed and marginally significant results for task ($F1,111=3.294$, $p = .07$), but no interaction. Duration was significantly longer for "Bob" when stressed as compared to the other unstressed words. The manner of elicitation (instruction or question) also affected duration of "Bob." Duration of "Bob" increased when following an instruction to stress the word, more increased when occurring in the natural context of responding to a question.

Total Duration

Analysis of the total duration of phrase length production indicated a statistically significant main effect for task ($F1,111=19.611$, $p <.001$). The total duration increased significantly when produced in response to a given instruction versus following the natural context of a question.

Variation within and between Productions

Further analysis of all measures was conducted to address the question of patterns of change for peak frequency, peak amplitude, and duration dependent on task. The differences for each variable were measured by comparing the peaks to the individual's average baseline for frequency, amplitude and

duration. These differences were then calculated within a contingency table. Findings indicated that when following a given instruction to stress a particular word, 95% of the productions involved an increase in peak frequency. In contrast, when producing a stressed word in response to the natural context of a question, only 80% of the productions involved increasing the peak frequency. All three parameters, frequency, amplitude and duration, were increased in 28.3% of the productions when directed to stress a word and in 25% when answering a question. Of interest, more variations occurred in the patterns of productions following the natural context of a question than that when an instruction was given (see figures 1 and 2).

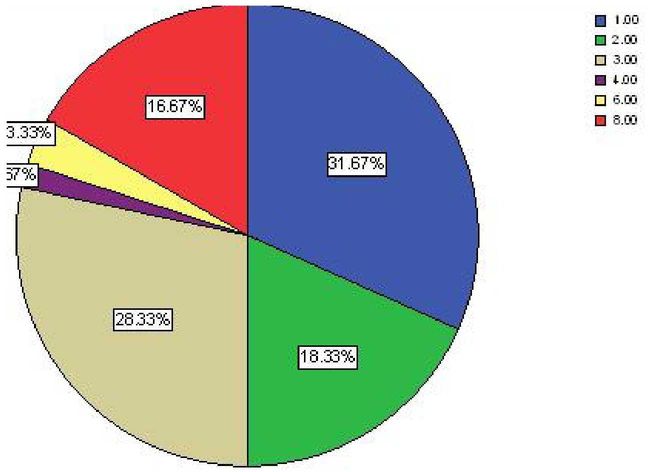

FIG. 1 RESPONSES FOLLOWING TASK INSTRUCTION

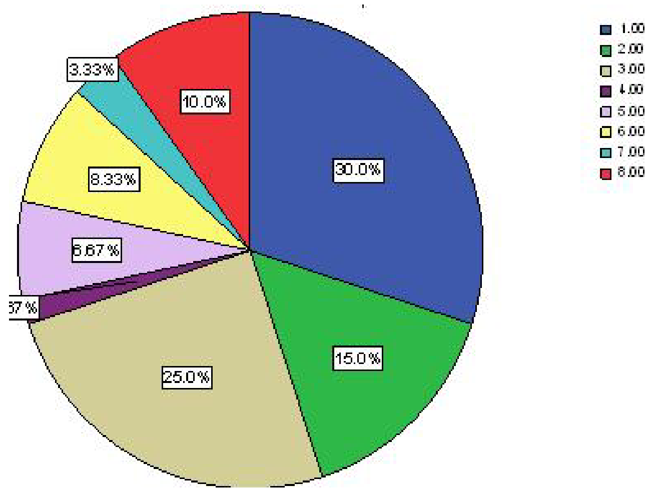

FIG. 2 RESPONSES IN NATURAL CONTEXT

Discussion

This study suggested that the acoustic parameters in correlation with stress are dependent upon the method used to elicit the production. A significant difference was found between the speech productions following elicitation of a production when given an instruction versus when following the natural context of a question. The stressed target words tended to have a higher mean peak frequency and amplitude and a longer duration than observed with the more natural productions. versus

For the purposes of this study, the acoustic parameters of peak frequency, peak amplitude and duration were measured for each word. A significant main effect for word was found in all conditions. This finding was expected in that it indicated that the stressed word was significantly different from the unstressed words in each of the productions. This finding demonstrated that the target word increased in magnitude for the acoustic parameters, regardless of task. Therefore, all productions were both quantitatively measured as stressed and were perceived as stressed. This increase in magnitude for stressed words has been well-documented in previous studies.

A significant main effect for task occurred in all measurements except for peak frequency of Bev and Bob. For these words, the performance was similar on these measures regardless of the task. This finding is in contrast to several studies in which a position effect for frequency has been reported. Previous studies have reported that fundamental frequency becomes declinated later in a phrase with structured tasks (Cooper et al., 1985). This study did not find a diminished effect for stress related to position.

English language is a polysyllabic language with diverse structure for syllables, syntax, and prosodic characteristics, whose natural diversity manifests within the natural variations that occur in speech. Lieberman (1960) indicated that typically two of the three acoustic parameters (frequency, amplitude, or duration) show an increase in magnitude during production of sentence stress. However, the Lieberman study used clearly enunciated and unambiguous sentences following a well-defined task and selection of best production. This study found that when following a directed task, productions followed a similar pattern as that reported by Lieberman (1960). An increase in magnitude for frequency or for all three parameters occurred on more than 60% of the productions. In contrast, following the more natural context of a question, the patterns of magnitude changes within the acoustic correlates of frequency, amplitude and duration were much more variable (see figures 1 and 2). By using a directed task and selecting the perceived best production, the natural variations observed in speech become reduced. The findings of this study have significant implications for both the

research methodology used to study the acoustic parameters of speech and for the ability to generalize the findings to natural speech.

Many studies have focused on frequency as the primary acoustic correlate for stress; while others have indicated that duration and intensity is a significant factor in determining stress (Cooper et al., 1985; Fry, 1958; Howell, 1993; Lieberman, 1960; Monnot, et al. 2003; Snow, 2000; Shtyrov, 2005). This study indicated that intensity should be considered with caution. The control of the acoustic parameters involved in stress production appears to be task dependent, with intensity being increased when a instruction was given. This finding may be related to a participant perception that increased loudness correlates with stress; however, in the more natural productions in response to a question, the use of intensity was not observed to the same extent.

Previous studies have indicated that duration is a significant factor when preparing clear speech or producing stress of a target word (Cooper et al., 1985; Ferguson, 2004; Liu, et al., 2004). While duration was found to increase significantly when participants were instructed to stress a particular word, duration did not have a similar magnitude of increase when responding to a question. Consideration should be taken into that duration may not play as much of a role in delineating stress in natural speech as it has been previously surmised from structured tasks.

Typically, previous studies analyzed "speech" using a directed task, with limited consideration for natural context. A problem arises in that the findings from these studies are generalized to natural speech and used to determine therapeutic interventions (Patel, 2003; Muerer et al., 2004). While the implications of these studies provide important information, the findings should be considered with caution when applied to populations. This study suggested that by using a directed task, the natural variations that occur within speech are reduced. The overall acoustic parameters of stress (frequency, amplitude, and duration) are not emphasized as much in a natural production, as we see with a directed task. An individual's perception of acoustic parameters of speech appears to influence production. Therefore, therapeutic intervention that is retraining prosodic skills through exaggerated speech to increase communicative intent may miss the important variation of natural productions by emphasizing less natural combinations of acoustic correlates of stress.

Conclusions

The production of acoustic correlates of stress is dependent on the task used and an individual's response to the task. When the participant is instructed in their production, the parameters of frequency, amplitude, duration, and total duration are significantly larger than those when responding to the more natural context of a question. Another difference is that the variations observed in natural speech are reduced with a directed task. Fewer combinations of the parameters were used when the participant was instructed to stress a particular word than those observed with answering a question. Further study is needed to examine the task oriented differences observed and the variations observed in natural speech. Implications from this study are that the research methodology used when studying the production of prosodic characteristics may need to be re-examined in the future. Understanding the ramifications of how a task effects the prosodic productions may affect future studies addressing the production, perception, and comprehension of prosody.

REFERENCES

Chen Y, Robb MP, Gilbert HR, and Lerman JW. (2001). A study of sentence stress production in Madarin speakers of American English. Journal of Acoustical Society of America. 109(4): 1681 – 1690.

Cooper WE, Eady SJ, and Mueller PR. (1985). Acoustical aspects of contrastive stress question-answer context. Journal of Acoustical Society of America. 77(6): 2142 – 2155.

Cutler A, Dahan D, and van Donselaar W. (1997). Prosody in comprehension of spoken language: a literature review. Language and Speech. 40: 141-201.

Eady, S.J. and Copper, W.E. (1986). Speech intonation and focus location in matched statements and questions. Journal of Acoustical Society of America. 80(2): 402- 413.

Ferguson SH. (2004). Talker differences in clear and conversational speech: vowel intelligibility for normal-hearing listeners. Journal of Acoustical Society of America.116 (4 Pt 1):2365-73.

Fry D. (1958). Experiments in the perception of stress.

Language and Speech. 1: 126-152.

Howell P. (1993). Cue trading in the production and perception of vowel stress. Journal of Acoustical Society of America. 94(4): 2063 – 2073.

Lieberman P. (1960). Some acoustic correlates of word stress in American English. Journal of Acoustical Society of America. 32: 451 -454.

Liu S, Rio ED, Bradlow AR, and Zeng F. (2004). Clear speech perception in acoustic and electric hearing. Journal of Acoustical Society of America. 116(4): 2374-2383.

Monnot M, Orbello D, Riccardo L, Sikka S, and Ross E. (2003). Acoustic analysis support subjective judgments of vocal emotion. Annals of New York Academy of Science. 1000: 288-292.

Muerer, E.M., Wender, M.C.O., von Eye Corleta, H., and Capp, E. (2004). Female suprasegmental speech parameters in reproductive age and menopause. Maturitas. 48: 71-77.

Munson B, Bjorum EM, and Windsor J. (2003). Acoustic and perceptual correlates of stress in nonwords produced by children with suspected developmental apraxia of speech and children with phonological disorder. Journal of Speech, Language and Hearing Research. 46: 189-202.

Patel R. (2003). Acoustic characteristics of the question-statement contrast in severe dysarthria due to cerebral palsy. Journal of Speech, Language and Hearing Research. 46(6):1401-15.

Pell M.D. (2007). Reduced sensitivity to prosodic attitudes in adults with focal right hemisphere brain damage. Brain and Language. 101(1):64 - 79.

Shtyrov Y, Pihko E, and Pulvermuller F. (2005). Determinants of dominance: is language laterality explained by physical or linguistic features of speech? NeuroImage. 27: 37-47.

Singer JD. (1998). Using SAS PROC MIXED to Fit multilevel models, hierarchical models, and individual growth models. Journal of Educational and Behavioral Statistics. 24(4): 323-5.

Snow D. (2000). The emotional basis of linguistic and nonlinguistic intonation: implications for hemispheric specialization. Developmental Neuropsychology. 17:1, 1 – 28.